

时序数据异常检测的可视化综述

韩东明¹ 左万利^{1,2} 彭涛^{1,2}

¹ (吉林大学计算机科学与技术学院 长春 130012)

² (符号计算与知识工程教育部重点实验室 (吉林大学) 长春 130012)

(liulu12@mails.jlu.edu.cn)

Tensor Representation Based Dynamic Outlier Detection Method in Heterogeneous Network

Liu Lu¹, Zuo Wanli^{1,2}, and Peng Tao^{1,2} **Name 五号**

¹ (College of Computer Science and Technology, Jilin University, Changchun 130012) **Depart. Correspond 小五号**

² (Key Laboratory of Symbol Computation and Knowledge Engineering (Jilin University), Ministry of Education, Changchun 130012)

Abstract Mining rich semantic information hidden in heterogeneous information network is an important task in data mining. The value, data distribution and generation mechanism of outliers are all different from that of normal data. It is of great significance of analyzing its generation mechanism or even eliminating outliers. Outlier detection in homogeneous information network has been studied and explored for a long time. However, few of them are aiming at dynamic outlier detection in heterogeneous networks. Many issues need to be settled. Due to the dynamics of the heterogeneous information network, normal data may become outliers over time. This paper proposes a dynamic Tensor Representation Based Outlier detection method, called TRBOutlier. It constructs tensor index tree according to the high order data represented by tensor. The features are added to direct item set and indirect item set respectively when searching the tensor index tree. Meanwhile, we describe a clustering method based on the correlation of short texts to judge whether the objects in datasets change their original clusters and then detect outliers dynamically. This model can keep the semantic relationship in heterogeneous networks as much as possible in the case of fully reducing the time and space complexity. The experimental results show that our proposed method can detect outliers dynamically in heterogeneous information network effectively and efficiently.

Abstract 五号, 至少 200 字, 影响 EI 索引

Key words dynamic outlier detection; heterogeneous information network; tensor representation; tensor index tree; clustering **Key words 五号**

摘要 挖掘隐藏在异质信息网络中丰富的语义信息是数据挖掘的重要任务之一。离群点在值、数据分布、和产生机制上都明显不同于正常数据对象。检测离群点并分析其不同的产生机制, 最终消除离群点具有重要的现实意义。目前, 针对异质信息网络动态离群点检测的研究工作相对较少, 还有很多问题有待解决。由于异质信息网络的动态性, 随着时间的变化, 正常数据对象也可能转变为离群点。针对异质网络提出一种基于张量表示的动态离群点检测方法, 并根据张量表示的高阶数据构建张量索引树。通过搜索张量索引树, 将特征加入到直接项集和间接项集中。同时, 根据基于短文本相关性的聚类方法来判断数据集中的数据对象是否偏离其原聚簇来动态检测网络中的离群点。该模型能够在充分降低时间和空间复杂度的条件下保留异质网络中的语义信息。实验结果表明, 该方法能够快速有效地进行异质网络环境下的动态离群点检测。

收稿日期: 2016-03-16; **修回日期:** 2016-04-26 **六号**

基金项目: 国家自然科学基金项目 (60903098); 吉林省工业技术研究和开发项目 (JF2012c016-2); 吉林大学研究生创新基金项目 (2015040)

This work is supported by the National Natural Science Foundation of China (60903098), the Project of Jilin Provincial Industrial Technology Research and Development (JF2012c016-2), and the Graduate Innovation Fund of Jilin University (2015040).

六号, 核实准确完整的基金名称和英文翻译

通信作者: 彭涛 (tpeng@jlu.edu.cn)

关键词 动态离群点检测; 异质信息网络; 张量表示; 张量索引树; 聚类

中图法分类号 TP391

时序数据是一系列基于一个准确时间测量的结果, 而它的时间间隔通常是规律的[1]。例如按照一定时间间隔统计得到的排名数据, 实时收集传输的传感器数据, 社交网络中用户每天的转发回复数据等生活中, 商业上, 军事里经常涉及的数据。

对于时序数据的可视分析在今天越来越广泛的应用在科学, 工程, 和商业领域中。数据可视化帮助人们利用感知减少认知负荷进而理解数据[2], 成功的被应用在不同环境下的时序数据分析中来[3]。例如社交媒体[4], 城市数据[5], 电子交易[6], 时序排名[7]。在不同领域的时序数据中, 发现重要的特征和趋势的日益增长的需求刺激了许多可视交互探索工具的发展 [8]: Line Graph Explore[9], LiveRAC[2], SignalLens[10]和 Data Vases[11]等。

时序数据的可视分析任务, 包括特征提取[14], 相关性分析和聚类[7], 模式识别[9], 异常检测[10]等。而异常检测在不同的研究领域都是一个重要的问题。异常检测表示发现数据中不符合预期行为的模式[12]。异常检测的目的是找到某些观察结果, 它与其他观察结果有很大的偏差, 很有可能是由于不同的原因或机制所产生的[17]。对应到不同的领域中, 网络安全中的异常表示网络设备异常或者可疑的网络状态[13]。情感分析中的异常表示一组数据中反常的观点, 情绪模式, 或者产生这些模式的特殊时机[16]。社交媒体中的异常可以是反常的行为, 例如网络机器人[20]。反常的传播过程, 例如谣言的传播[19]。这些异常信息或模式的产生, 例如电脑侵入, 社交机器人, 道路拥堵状况等。都会极大影响日常生活, 社会稳定, 国家发展。提早发现识别这些异常有助于及时认清实际状况, 找出产生原因, 从而进一步分析解决问题。

异常检测在如今已经有许多成熟的方法, 而且在机器学习领域也引起了广泛的关注[12], 包括有监督[21]和无监督的异常检测方法[22]。自动化的学习算法通常基于以下假设, 即有充足的训练数据可用, 同时这些数据反应的是正常的行为。否则, 很有可能因为新的观测数据是不常见的正常事件[25], 而导致了错误的分类。但当涉及到人工标注数据的问题时, 往往需要大量的数据, 费事费力, 难以获取, 同时又十分依赖于主观人为的判断, 这些因素极大地影响了最后分析结果的质量[20]。与此同时, 如何在自然数据中定义其中正常或异常的行为也是十分困难的[23]。此时人类的经验和知识在异常检测方面便显现出了优势, 它可以被用来更新改进模型, 已及对异常检测

过程进行实时控制。

如今的大数据时代, 面对数据维度多, 数据尺寸大的场景, 具有强有力的功能和巨大的价值的数据可视化在其中可以更好的帮助人来分析理解数据和其中的行为, 模式等。异常检测方法, 与人的经验和背景知识, 以及交互式可视化的灵活思维相结合, 可以帮助人们发现从未想到的异常, 减少人的劳动, 提高异常的检测识别能力。

1 挑战

异常检测的挑战在 Chandola V. [12]等人的综述中, 已经进行了全面的总结。异常检测首先要先定义异常, 而正常和异常的界限往往是难以区分的, 特别是当界限被规定后, 在界限附近的观测值, 很容易把正常当做异常, 亦或是把正常当做异常, 特别是一些设定阈值的异常检测方法, 很难处理此种问题, 需要其他的信息进行辅助判断。以下为异常检测的 5 种挑战。

- (a) 有些异常的行为通常是人为恶意操控的, 它会模仿现实中的真实正常的行为, 让异常的现象观测起来和正常现象一样, 导致异常检测的任务变得十分困难。例如社交网络中机器人[20]回复, 它会模仿真实人类的语气, 时间频率等特征, 以假乱真。
- (b) 随着发展和进步, 许多领域的正常行为也在与时俱进, 其概念可能在未来会失效。而且很多数据集都是复杂和动态的, 例如传感器数据[26][25][42], 网络安全数据[45]等, 这些挑战在文献[19]中都有提到。而在复杂多变时效性很高的场景中, 需要人的监督和判断来进行异常的检测及分析。
- (c) 领域间的技术很难互相应用, 不同领域间的实际情况不一样, 有些异常的现象在其它领域可能就是正常的情况。
- (d) 用于训练确认异常的模型所使用的标记数据十分难以获取。
- (e) 数据中包含的噪音和异常往往很相似, 如何去区分和清晰也是面临的挑战之一。

Table 1 Classification of anomaly and relevant visualization tasks with Visualization forms

表 1 异常分类和相关的可视化任务以及可视化组成

Type	Papers	Data
Attribute	[15] [18] [23]	Social media
	[24] [45]	Network traffic
	[35]	Business data
	[25]	Sensor data
	[41] [52]	Traffic data
Topology	[26]	Sensor data
	[56]	Network traffic
	[27]	Firewall policy
Hybridize	[19][20]	Social media
	[34]	network management

2 异常分类

时序数据在不同的环境和应用领域中, 会包含许多的领域相关的信息及属性。例如网络数据中流量节点的类型, 社交媒体上个人的信息[19], 注册时间等信息都可以视为时序数据中的属性。时序数据的异常情况便可以通过附属在时间维度的属性来进行分析检测。除此之外, 数据中实体间蕴含的关系也会揭示时序上的异常情况, 例如社交网络中信息转发回复的会话网络[19], 动态图的网络演变等数据和场景中都包含了实体间的关系, 即拓扑结构。

时序数据异常检测可视化中, 可基于拓扑结构或属性来对异常进行分析检测。下面将从三个角度去对已有的时序数据异常检测可视化工作进行分类。分别是属性上, 拓扑结构上, 和混合情况。例如时空数据中的地理信息, 传感器网络的传输顺序, 网络数据中的节点类型, 动态图中的拓扑结构。属性和拓扑结构在时序上的变化多端, 规律难寻。而通过可视化的方法来探究时序数据中的异常, 与人的知识和经验相结合, 将会有事半功倍的效果。

2.1 属性

本节介绍基于时序数据中的属性进行异常检测的相关可视化工作, 主要是对时序上属性变化的异常进行可视分析。

Dennis Thom[15]等人的工作中根据信息的内容和发送的地理位置等属性进行聚类, 形成标签云, 用来可视分析时空数据中异常情况, 例如地震, 骚动等。Schreck T[18]等人的工作则是将实时数据和历史数据进行对比, 包括频率, 内容等多种属性, 探寻时序上的属性异常。文献[24]对网络流量异常进行检测, 基于不同时间内的属性例如 DNS, HTTP 等请求数目或

流量, 检测相对应的异常情况, 例如 DNS 攻击, 探测攻击等。文献[35]对运营数据中的异常流程进行分析, 如图 2.(c)所示, 通过在弦图中可视化每两个实例之间的关系, 设定阈值以及通过判定关系之间的交叉和时序上的变化, 进行异常如诈骗的检测。Voila[41]系统提出了一种基于张量的异常分析算法用于转换时空数据, 如出租车行驶数据中时序变化的交通流量, 位置等信息, 如图 1.(a)所示, 得到的张量时间序列再去结合历史数据进行期望模式分析, 之后对应张量分解后的上下文进行异常检测, 同时可以根据用户的交互对异常的模式进行排序。文献[25]基于传感器数据来检测船只的异常状况, 例如船的速度值异常, 船舶靠近海岸线等历史数据中匹配不到的异常行为。

2.2 拓扑

本节介绍基于时序数据中的拓扑结构进行异常检测的相关可视化工作, 即原始数据中存在拓扑结构, 或者把数据抽象成拓扑结构, 进而对拓扑结构在时序上的异常进行可视分析。

SAVE[26]是一个用于检测传感器数据异常变化的时序拓展模型(TEM), 在拓扑视图中如图 2.(a)所示, 把传感器节点置于环形布局上, 用颜色表示时间的先后顺序, 可以观测传感器时序上拓扑结构的变化, 进而发现异常。PolicyVis[27]系统用于防火墙安全策略的可视监测, 巧妙的把安全策略的规则命令的逻辑顺序转化成拓扑结构, 不同的异常对应不同的拓扑结构, 按照时间和流量等维度绘制不同规则下所代表的矩形, 矩形间的重合所带来的阴影则代表了不同的情况, 其中就包含如策略冲突等异常情况。

文献[56]设计一种新型的层次结构和节点连接相结合的可视化方法, 如图 2.(b)中所示。系统可以用来检测网络流量中异常, 并灵活的创建相关维度属性

之间的响应。

2.3 混合

本节介绍结合时序数据中的拓扑结构和属性, 进行异常检测的相关可视化工作,

Twitter 上消息的回复转发构成了会话网络, FluxFlow[19]系统用于对会话传播过程进行异常检测, 探索, 解释。它的分析模块结合了多种机器学习算法组来探索异常转发的特征。例如用户注册的时间, 朋友数量等属性上的特征。图 1.(b)设计了一个时序上的包裹圆, 用于展示原始信息随着时间的推移在用户之间传播的可视化视图, 并通过 OCCRF 模型[28]

来计算其中每一个圆(用户)的异常分数, 其中圆的大小编码了用户的重要性。结合拓扑结构和属性上的特征进行异常检测。TargetVue[20], 如图 1.(c)所示, 用于检测社交网络中具有异常行为的用户, 系统最初为每个用户提取一组行为特征, 并使用异常检测算法来发现可疑的用户。最可疑的用户将会用两种类型的图标进行可视化编码, 一种编码了他们的通信行为(即发布消息, 转发消息); 另一种编码了与之相应的行为的特征。Qi Liao[34]等人的工作中研究管理网络数据中的动态异常问题, 从多个层面和角度去关注节点, 关系, 和社团在时序上的异常变化。

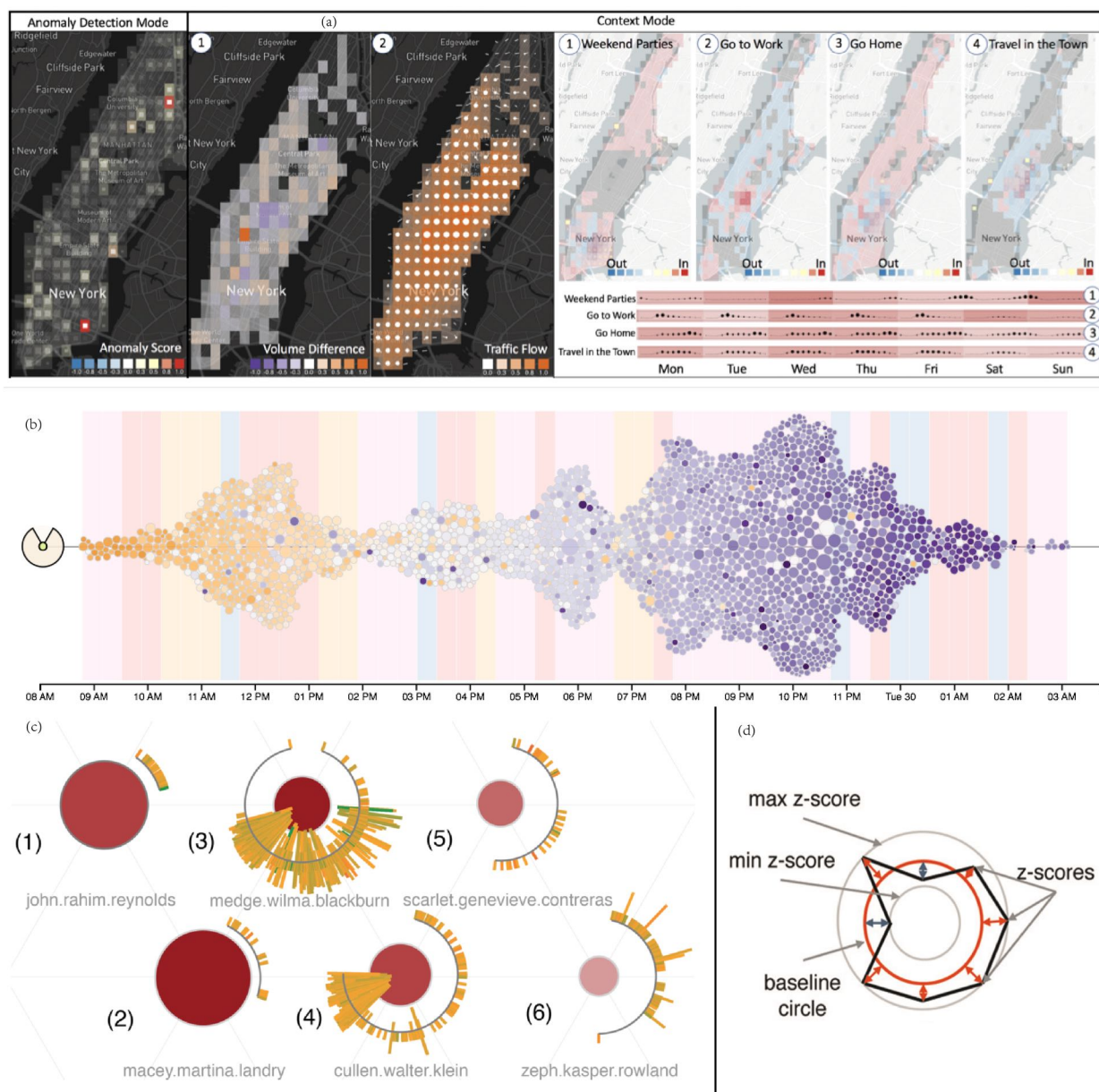


Fig. 1 The tensor index tree constructed by a third order tensor in bibliographic network.

图 1 文献网络中由 3 阶张量所构建的张量索引树

Table 2 Classification of anomaly detection method and relevant visualization tasks with Visualization forms

表 2 异常检测方法分类和相关的可视化任务以及可视化组成

Type	Papers	Data	
Feature	[23] [44]	Social media	
	[29],[41],[40], [39][24]	Network traffic	
	[48]	Spatiotemporal	
Cluster	[15]	Social media	
	[35]	Business data	
	[54]	Temperature data	
	[53]	Traffic data	
	[51][55]	/	
Machine learning	[25]	Maritime data	
	[19] [20]	Social media	
	[18]	Social media	
Others	[27]	Firewall policy	
	[45]	Network traffic	

3 异常检测方法分类

本节根据异常检测的方法，对已有的时序数据异常检测可视化的工作进行分类。异常检测方法在[49][50][12]中被介绍了许多，针对于不同的领域和不同的类型数据，异常检测方法也不尽相同，同一种异常检测方法，在面对不同实际情况时，效果可能会大打折扣。结合异常检测算法，加上可视分析方法和人类的知识经验的结合，可以更加精确，多角度的检测出以前难以单独用异常方法检测到的异常。下面通过直接投影方法，聚类，机器学习三个方面对时序数据异常检测的可视化工作进行分类。

3.1 直接投影方法

Celenk M,[29]基于短期的网络特征和平均时间熵的观测结果，针对对异常网络流量设计了 FLD 图，用于对网络异常中的统计特征进行可视分析。Z-Glyph[23]被设计用于检测多元数据中的离群点，如图 1.(d)所示，可以配合多个统一视图(Small Multiples)进行时序上异常检测。同样检测离群点的方还有，SOM[32]方法，用平行坐标轴[31]的方法，以及盒须图[33]组件。文献[48]设计了与时空属性相结合，基于仿真数据的盒须图，用于对例如环境降雨，全球环流，海面温度等数据进行异常检测。

文献[24]用于网络流量进行异常检测，对不同时间内的不同类型流量进行可视化，投影到六边形视图上。每种异常都会有特定的特征在六边形视图中显示，例如 DoS 攻击，探测攻击等异常。

2004 年[38]PCA 第一次被提出用于流量异常的

检测，之后 Brauckhoff D 等人[37]提出了一个解决方案用于处理数据中时间相关性的问题，让 PCA 可以更好的应用于异常检测。此外还有基于直方图[40]，最大熵估计[39]等方法用于此项分析任务，PCA 在时空数据的异常检测中经常被用到，但是数据中如果包含张量结构，PCA 方法可能会隐藏一些有特殊意义的异常，综述错误! 未找到引用源。中介绍了基于张量来检测识别异常的新技术手段。文献 [44]用三阶张量来表示动态变化的社交网络，并从中检测异常的演变。[41]用张量来表示时空数据，根据动态的张量网络来结合多个统一视图(Small Multiples)和图标(Glyph)来可视化车流交通的异常情况，并通过张量分解从数据中推测出异常的特征和模式。

3.2 聚类方法

聚类用来对实体进行相似性分组的过程，或者用于大数据的缩减技术。

文献[15]构建了适用于分析来自连续数据流，例如微博帖子的增强 Lloyd 聚类方法，系统包括一个用于探索捕获重要事件及异常的工分析工作台和针对时空数据的异常检测模块，并在地图视图上根据位置去布局聚类好的标签。

文献[51]使用了结合图可视化的最小生成树聚类方法，在一些时序数据的线性回归中进行异常检测，并提出多种用于分类异常的图形特征。文献[53]中用 X-means 聚类算法并与可视化相结合，用于对时空交通数据中流量变化，进行包括异常检测在内的多种任务分析。文献[54]提出一种基于折线图的三维时变可视化技术，用于提取显示相似的值，并应用

SAX((Symbolic Aggregate approXimation)来检测频繁或异常的模式。文献[55]也是通过聚类一组时序变化的值来搜索其中的异常值。文献[35]使用多个统一视图以及弦图中可视化显示对于商业管理数据多个维度中的影响因子,并运用相关性分析,局部匹配以及聚类的技术来提取其中重要的影响因素。

3.3 机器学习方法

SOM[46]可以看做是一个基于神经网络的聚类算法,高维数据投影到 2D 空间后可以识别定义不同聚类间的边界[47],许多工作都会加入一个高位投影视图用于数据集分类的总览。然而对于很难明确的对数据进行分类的情况时,SOM 并不能提供一个效果很好的解决方案时,GMM 便可以应用于 SOM 上。

文献[25]根据历史的观察数据建立正常的模型,之后对新生的实时数据进行异常行为检测,如果计算的累计概率值高于设定阈值便通知用户进行判断,来对模型进行迭代。其中的训练器和检查器基于 GMM

和 SOM 完成,一个控制视图用于过滤设定参数,例如传感器的选取,覆盖范围过滤,累计的概率值阈值等。概览地图上标注船舶的类型速度等属性,当发生异常的时候会在地图上进行标注,人为判断是否是异常情况,以便对模型进行迭代。

文献[19][20]中都基于 OCCRF 模型对 Twitter 数据中的异常转发等用户行为进行异常检测,并结合多种机器学习算法探索社交网络转过程的重要特征。

3.4 其他

还有一些工作把数据通过一些例如图标法,地理视图,矩阵图等可视化方法,用于对异常的数据进行检测。通过颜色的差异[27],地理上视图上[18]的高亮,多种统计图相关的方法来显示数据中的异常。例如文献[18]将实时数据和历史数据的内容及频率等属性进行对比,在地理视图上对异常的属性信息进行高亮。

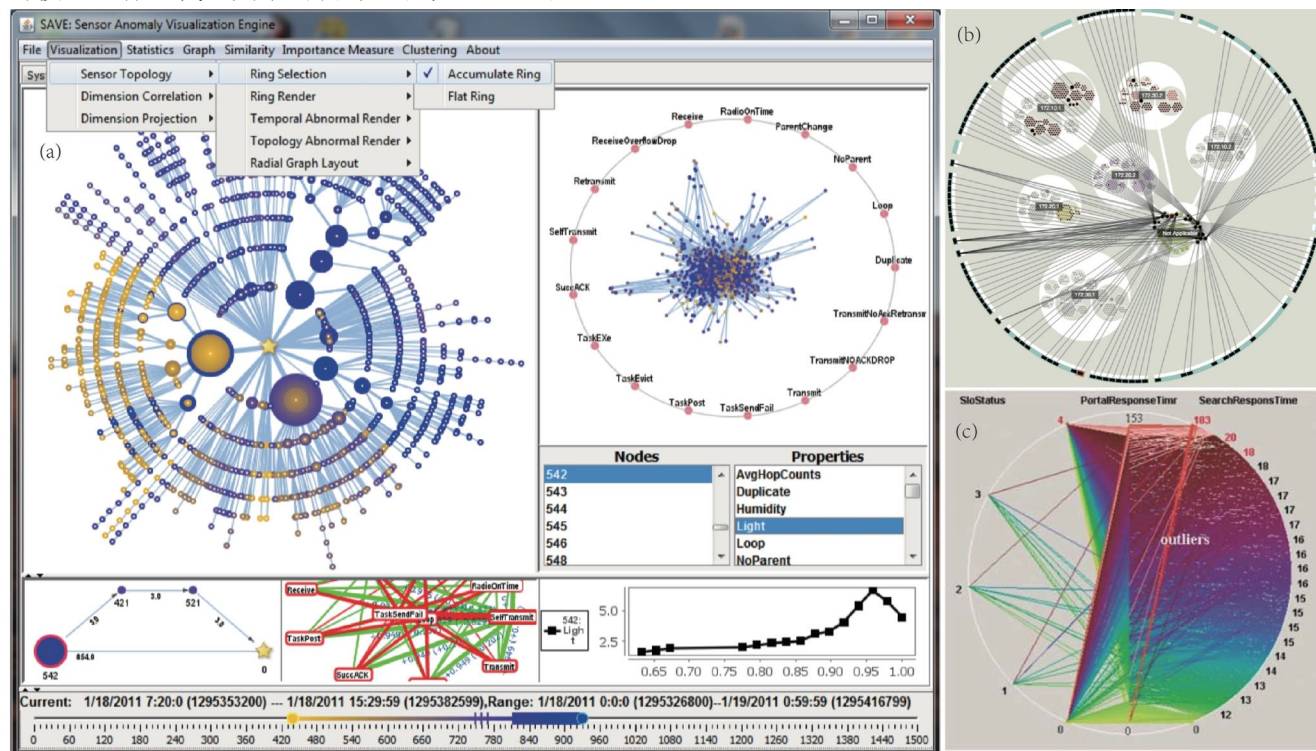


Fig. 2 The tensor index tree constructed by a third order tensor in bibliographic network.

图 2 文献网络中由 3 阶张量所构建的张量索引树

4 总结和未来发展

本文对于时序数据的异常检测可视化工作进行了综述,并针对现有的工作,对时序数据的异常类型和异常检测分别进行了分类。近年来,随着可视化方法在时序数据异常检测上的不断应用,逐步展现出可视化的巨大优势,例如社交领域[36],现有的自动化

方法的内在局限性,通过可视分析来检测异常的用户行为是一个十分有前途的方向,许多视觉分析系统通过用户专业的知识和经验,但是其中的挑战,例如不同领域的迁移,随时代变化的异常的判定标准和类型等,都将成为该研究方向在未来需要攻克的问题。

异常检测任务特别的需要先验知识,即人的经验和背景知识来进行分析判断,可视化可以更好的对数据进行抽象,表达,发现自动算法不能判别的异常

情况。自动的异常检测方法,人的知识经验和交互式可视分析方法,三者相互结合,可以准确的判断异常,全面的发现异常,更智能的检测异常,会给人们的生活,国家的进行,社会的发展带来了巨大的安全保障和快速发展。

参 考 文 献

详细格式要求参照期刊主页下载中心“参考文献规范”,

<http://crad.ict.ac.cn/CN/column/column33.shtml>

- [1] Bojan V C, Raducu I G, Pop F, et al. Cloud-based service for time series analysis and visualisation in Farm Management System[C]//Intelligent Computer Communication and Processing (ICCP), 2015 IEEE International Conference on. IEEE, 2015: 425-432.
- [2] McLachlan P, Munzner T, Koutsosios E, et al. LiveRAC: interactive visual exploration of system management time-series data[C]//Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. ACM, 2008: 1483-1492.
- [3] Weber M, Alexa M, Müller W. Visualizing time-series on spirals[C]//Infovis. 2001, 1: 7-14.
- [4] Kumar P, Sinha A. Real-time analysis and visualization of online social media dynamics[C]//Next Generation Computing Technologies (NGCT), 2016 2nd International Conference on. IEEE, 2016: 362-367.
- [5] Chen W, Huang Z, Wu F, et al. VAUD: A Visual Analysis Approach for Exploring Spatio-Temporal Urban Data[J]. IEEE Transactions on Visualization & Computer Graphics, 2017 (1): 1-1.
- [6] Xie C, Chen W, Huang X, et al. Vaet: A visual analytics approach for e-transactions time-series[J]. IEEE transactions on visualization and computer graphics, 2014, 20(12): 1743-1752.
- [7] Xia J, Hou Y, Chen Y V, et al. Visualizing Rank Time Series of Wikipedia Top-Viewed Pages[J]. IEEE Computer Graphics and Applications, 2017, 37(2): 42-53.
- [8] Cho M, Kim B, Bae H J, et al. Stroscope: Multi-scale visualization of irregularly measured time-series data[J]. IEEE transactions on visualization and computer graphics, 2014, 20(5): 808-821.
- [9] Kincaid R, Lam H. Line graph explorer: scalable display of line graphs using focus+ context[C]//Proceedings of the working conference on Advanced visual interfaces. ACM, 2006: 404-411.
- [10] Kincaid R. Signallens: Focus+ context applied to electronic time series[J]. IEEE Transactions on Visualization and Computer Graphics, 2010, 16(6): 900-907.
- [11] Thakur S, Rhyne T M. Data vases: 2d and 3d plots for visualizing multiple time series[J]. Advances in Visual Computing, 2009: 929-938.
- [12] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. ACM computing surveys (CSUR), 2009, 41(3): 15.
- [13] Pearlman J, Rheingans P. Visualizing network security events using compound glyphs from a service-oriented perspective[M]//VizSEC 2007. Springer, Berlin, Heidelberg, 2008: 131-146.
- [14] Alonso O, Khandelwal K. Kondenser: Exploration and visualization of archived social media[C]//Data Engineering (ICDE), 2014 IEEE 30th International Conference on. IEEE, 2014: 1202-1205.
- [15] Thom D, Bosch H, Koch S, et al. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages[C]//Pacific visualization symposium (PacificVis), 2012 IEEE. IEEE, 2012: 41-48.
- [16] Wang Z, Joo V, Tong C, et al. Anomaly Detection through Enhanced Sentiment Analysis on Social Media Data[C]//Cloud Computing Technology and Science (CloudCom), 2014 IEEE 6th International Conference on. IEEE, 2014: 917-922.
- [17] Hawkins D M. Identification of outliers[M]. London: Chapman and Hall, 1980.
- [18] Schreck T, Keim D. Visual analysis of social media data[J]. Computer, 2013, 46(5): 68-75.
- [19] Zhao J, Cao N, Wen Z, et al. # FluxFlow: Visual analysis of anomalous information spreading on social media[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1773-1782.
- [20] Cao N, Shi C, Lin S, et al. Targetvue: Visual analysis of anomalous user behaviors in online communication systems[J]. IEEE transactions on visualization and computer graphics, 2016, 22(1): 280-289.
- [21] Steinwart I, Hush D, Scovel C. A classification framework for anomaly detection[J]. Journal of Machine Learning Research, 2005, 6(Feb): 211-232.
- [22] Eskin E, Arnold A, Prerau M, et al. A geometric framework for unsupervised anomaly detection: Detecting intrusions in unlabeled data[J]. Applications of data mining in computer security, 2002, 6: 77-102.
- [23] Cao N, Lin Y R, Gotz D, et al. Z-Glyph: Visualizing outliers in multivariate data[J]. Information Visualization, 2017: 1473871616686635.
- [24] Onut I V, Zhu B, Ghorbani A A. A novel visualization technique for network anomaly detection[C]//PST. 2004: 167-174.
- [25] Riveiro M, Falkman G, Ziemke T. Improving μ anomaly detection and situation awareness through interactive visualization[C]//Information Fusion, 2008 11th International Conference on. IEEE, 2008: 1-8.
- [26] Shi L, Liao Q, He Y, et al. SAVE: Sensor anomaly visualization engine[C]//Visual Analytics Science and Technology (VAST), 2011 IEEE Conference on. IEEE, 2011: 201-210.
- [27] Tran T, Al-Shaer E S, Boutaba R. PolicyVis: Firewall Security Policy Visualization and Inspection[C]//LISA. 2007, 7: 1-16.
- [28] Song Y, Wen Z, Lin C Y, et al. One-Class Conditional Random Fields for Sequential Anomaly Detection[C]//IJCAI. 2013: 1685-1691.
- [29] Celenk M, Conley T, Willis J, et al. Predictive network anomaly detection and visualization[J]. IEEE Transactions on Information Forensics and Security, 2010, 5(2): 288-299.
- [30] Kandogan E. Visualizing multi-dimensional clusters, trends, and outliers using star coordinates[C]//Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. ACM,

- 2001: 107-116.
- [31] Novotny M, Hauser H. Outlier-preserving focus+ context visualization in parallel coordinates[J]. IEEE Transactions on Visualization and Computer Graphics, 2006, 12(5): 893-900.
- [32] Munoz A, Muruzabal J. Self-organizing maps for outlier detection[J]. Neurocomputing, 1998, 18(1): 33-60.
- [33] Kampstra P. Beanplot: A boxplot alternative for visual comparison of distributions[J]. 2008.
- [34] Liao Q, Striegel A. Intelligent network management using graph differential anomaly visualization[C]//Network Operations and Management Symposium (NOMS), 2012 IEEE. IEEE, 2012: 1008-1014.
- [35] Hao M C, Keim D A, Dayal U, et al. Business process impact visualization and anomaly detection[J]. Information Visualization, 2006, 5(1): 15-27.
- [36] Wu Y, Cao N, Gotz D, et al. A survey on visual analytics of social media data[J]. IEEE Transactions on Multimedia, 2016, 18(11): 2135-2148.
- [37] Brauckhoff D, Salamatian K, May M. Applying PCA for traffic anomaly detection: Problems and solutions[C]//INFOCOM 2009, IEEE. IEEE, 2009: 2866-2870.
- [38] Lakhina A, Crovella M, Diot C. Diagnosing network-wide traffic anomalies[C]//ACM SIGCOMM Computer Communication Review. ACM, 2004, 34(4): 219-230.
- [39] Gu Y, McCallum A, Towsley D. Detecting anomalies in network traffic using maximum entropy estimation[C]//Proceedings of the 5th ACM SIGCOMM conference on Internet Measurement. USENIX Association, 2005: 32-32.
- [40] Kind A, Stoecklin M P, Dimitropoulos X. Histogram-based traffic anomaly detection[J]. IEEE Transactions on Network and Service Management, 2009, 6(2).
- [41] Cao N, Lin C, Zhu Q, et al. Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data[J]. IEEE transactions on visualization and computer graphics, 2018, 24(1): 23-33.
- [42] Wang X, Lizier J, Obst O, et al. Spatiotemporal anomaly detection in gas monitoring sensor networks[J]. Wireless Sensor Networks, 2008: 90-105. MLA
- [43] Gama J, Tork H F. Tensor-based anomaly detection: An interdisciplinary survey[J]. 2016.
- [44] Dereszynski E W, Dietterich T G. Spatiotemporal models for data-anomaly detection in dynamic environmental monitoring campaigns[J]. ACM Transactions on Sensor Networks (TOSN), 2011, 8(1): 3.
- [45] McKenna S, Staheli D, Fulcher C, et al. Bubblenet: A cyber security dashboard for visualizing patterns[C]//Computer Graphics Forum. 2016, 35(3): 281-290.
- [46] Kohonen T. The self-organizing map[J]. Neurocomputing, 1998, 21(1): 1-6.
- [47] Kraiman J B, Arouh S L, Webb M L. Automated anomaly detection processor[J]. Proceedings of SPIE: Enabling Technologies for Simulation Science VI, 2002, 4716: 128-137.
- [48] Sun Y, Genton M G. Adjusted functional boxplots for spatio-temporal data visualization and outlier detection[J]. Environmetrics, 2012, 23(1): 54-64.
- [49] Hodge V, Austin J. A survey of outlier detection methodologies[J]. Artificial intelligence review, 2004, 22(2): 85-126.
- [50] Zhang T, Wang X, Li Z, et al. A survey of network anomaly visualization[J]. Science China Information Sciences, 2017, 60(12): 121101.
- [51] Kim S S, Krzanowski W J. Detecting multiple outliers in linear regression using a cluster method combined with graphical visualization[J]. Computational Statistics, 2007, 22(1): 109-119.
- [52] Shekhar S, Lu C T, Liu R, et al. CubeView: a system for traffic data visualization[C]//Intelligent Transportation Systems, 2002. Proceedings. The IEEE 5th International Conference On. IEEE, 2002: 674-678.
- [53] Malinao J A, Juayong R A B, Becerra J G, et al. Patterns and Outlier Analysis of Traffic Flow Using Data Signatures via IDIRBrG Method and Vector Fusion Visualization[C]//Human-Centric Computing (HumanCom), 2010 3rd International Conference on. IEEE, 2010: 1-6.
- [54] Imoto M, Itoh T. A 3d visualization technique for large scale time-varying data[C]//Information Visualisation (IV), 2010 14th International Conference. IEEE, 2010: 17-22.
- [55] Lin J, Keogh E, Lonardi S. Visualizing and discovering non-trivial patterns in large time series databases[J]. Information visualization, 2005, 4(2): 61-82.
- [56] Arendt D L, Burtner R, Best D M, et al. Ocelot: user-centered design of a decision support visualization for network quarantine[C]//Visualization for Cyber Security (VizSec), 2015 IEEE Symposium on. IEEE, 2015: 1-8.



Liu Lu, born in 1989. PhD. Her main research interests include Web mining, information retrieval, machine learning.



Zuo Wanli, born in 1957. PhD, professor, PhD supervisor. Senior member of China Computer Federation. His main research interests include database theory, data mining, Web mining, machine learning, and Web search engine.



Peng Tao, born in 1977. PhD, professor. Member of China Computer Federation. His main research interests include Web mining, information retrieval, and machine learning.

作者介绍小五号，照片最好正面免冠，不要侧脸照，照片背景尽量简单。作者介绍主要包括：姓名、出生年月、学历、职称、头衔和研究领域。